Applying Cyberinfrastructure and Geospatial Semantics to Advance Discoveries in GIScience

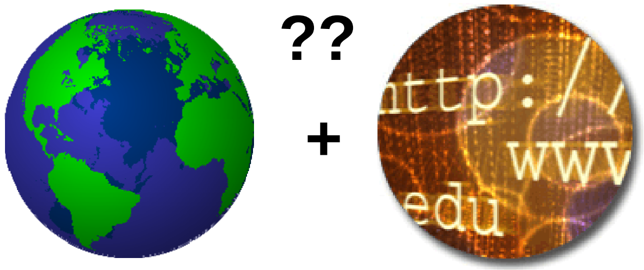
Wenwen Li
Center for Spatial Studies &
Department of Geography
University of California, Santa Barbara

04-16-2012

UCSB

Outline

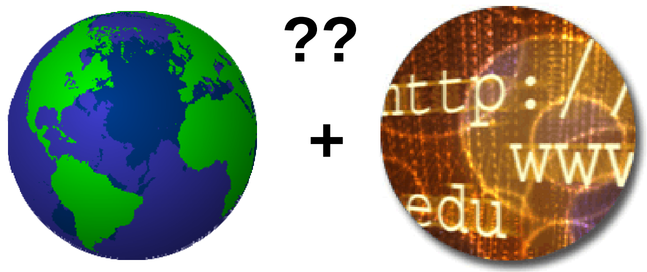
- Motivation
- State-of-the-art Solutions
  - Part I: Sharing of geoscientific data – OGC
  - Part II: A hybrid approach for service discovery
  - Part III: Service integration and chaining
  - Part IV: Visualization
  - Part V: Demo
- Future Work



UCSB

Outline

- Motivation
- State-of-the-art Solutions
  - Part I: Sharing of geoscientific data – OGC
  - Part II: A hybrid approach for service discovery
  - Part III: Service integration and chaining
  - Part IV: Visualization
  - Part V: Demo
- Future Work



A few questions

- Have you ever experienced a hard time looking for data?
- Have you ever asked that “why isn’t an easy-to-use tool to solve my problem?”
- Have you complaint about “why it takes so long to get the data processing done?”

UCSB

Problem1 – Data discovery

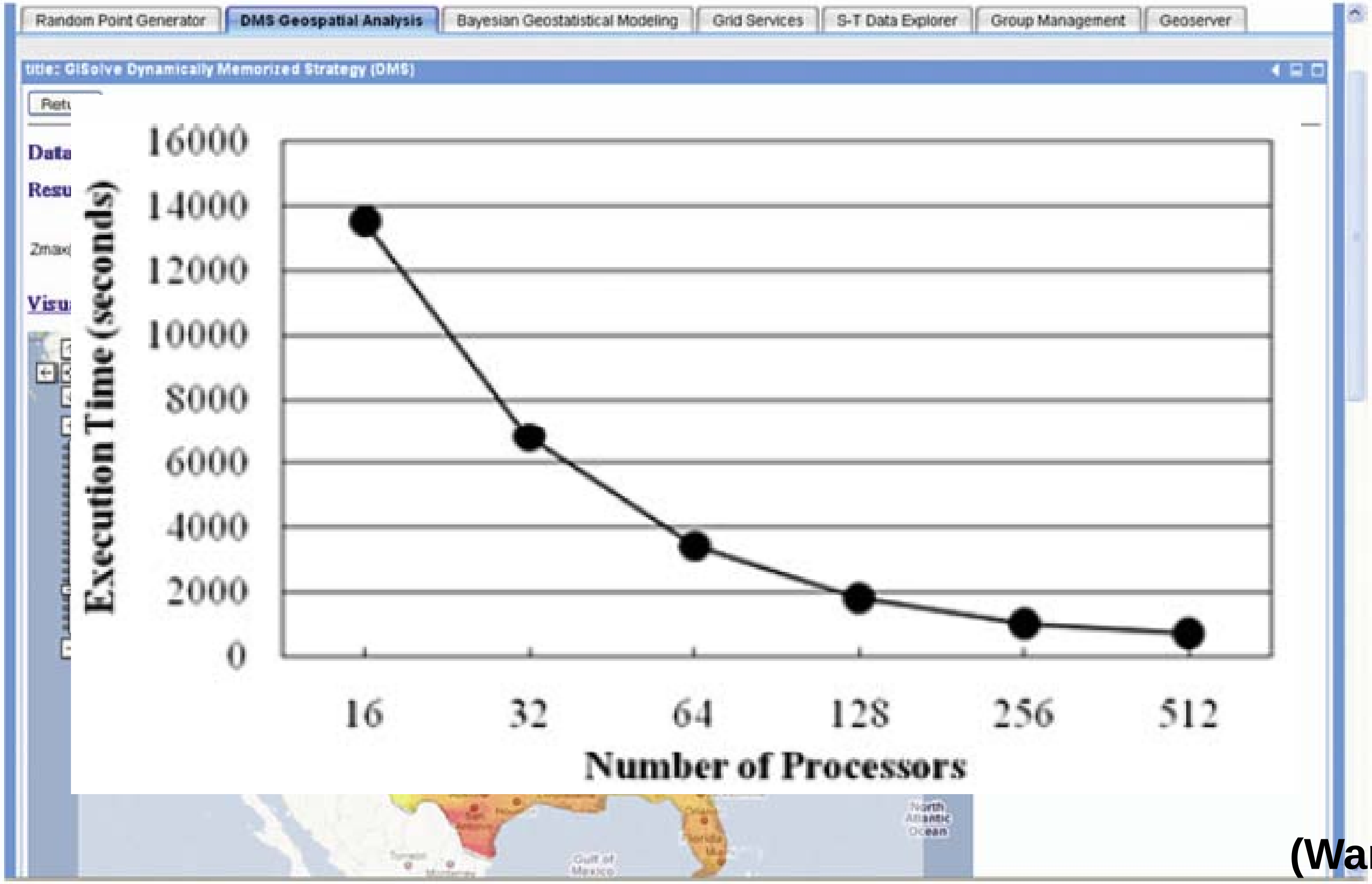
- **Web explosion**
  - 11.6 billion of webpages on WWW (2005)
- **Development of Earth observation technique**
  - EOSDIS 3TB daily
  - Earthscope 67TB, double every 2 months
- **Personal computer**
  - Hard-drive: 20G → 1TB

Problem 2 – GIS Tool

- **Data heterogeneity**
  - Different sources:
    - NASA. USGS, USCB; ESRI, Individual
  - Different formats:
    - ESRI Shape, Coverage, TIFF, GPS
- **Multiple tools available**
  - Example: Web Service
    - ESRI ArcGIS Server
    - MN MapServer
    - GeoServer
  - Example: Batch geocoding
  - Integration: one tool cannot satisfy a GIS task

UCSB

Problem3- Efficiency of GIS Data Processing



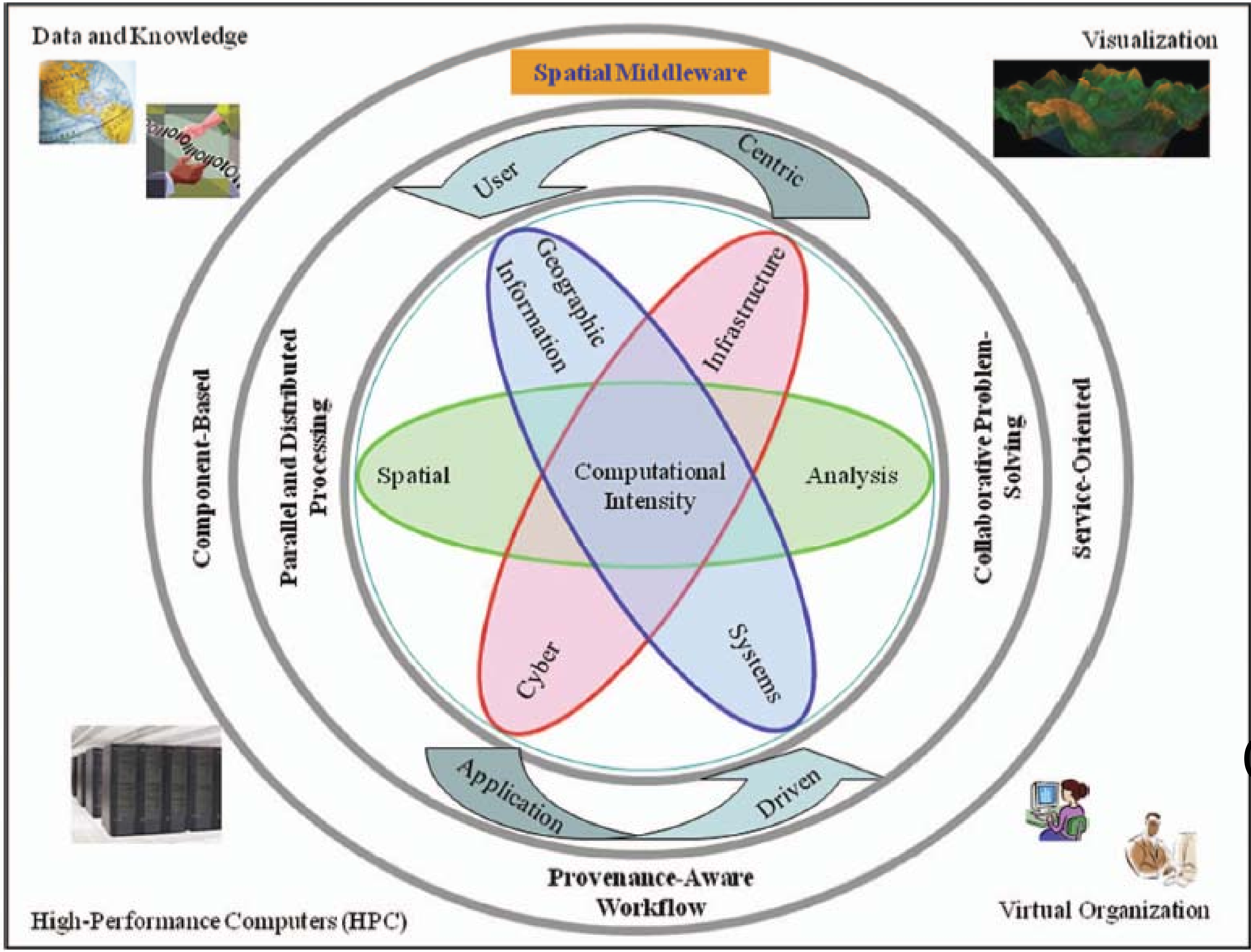
(Wang, 2010)

Daily data: 4 hours – one day's data -→ 2 months – one year's data

What is Cyberinfrastructure?

- **“Cyberinfrastructure** is the coordinated aggregate of software, hardware and other technologies, as well as human expertise, required to support current and future discoveries in science and engineering. The challenge of Cyberinfrastructure is to integrate relevant and often disparate resources to provide a useful, usable, and enabling framework for research and discovery characterized by broad access and “end-to-end” coordination”
- **Cyberinfrastructure** consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible.

What is Cyberinfrastructure?



(Wang, 2010)

UCSB

What's Geospatial Semantic Web

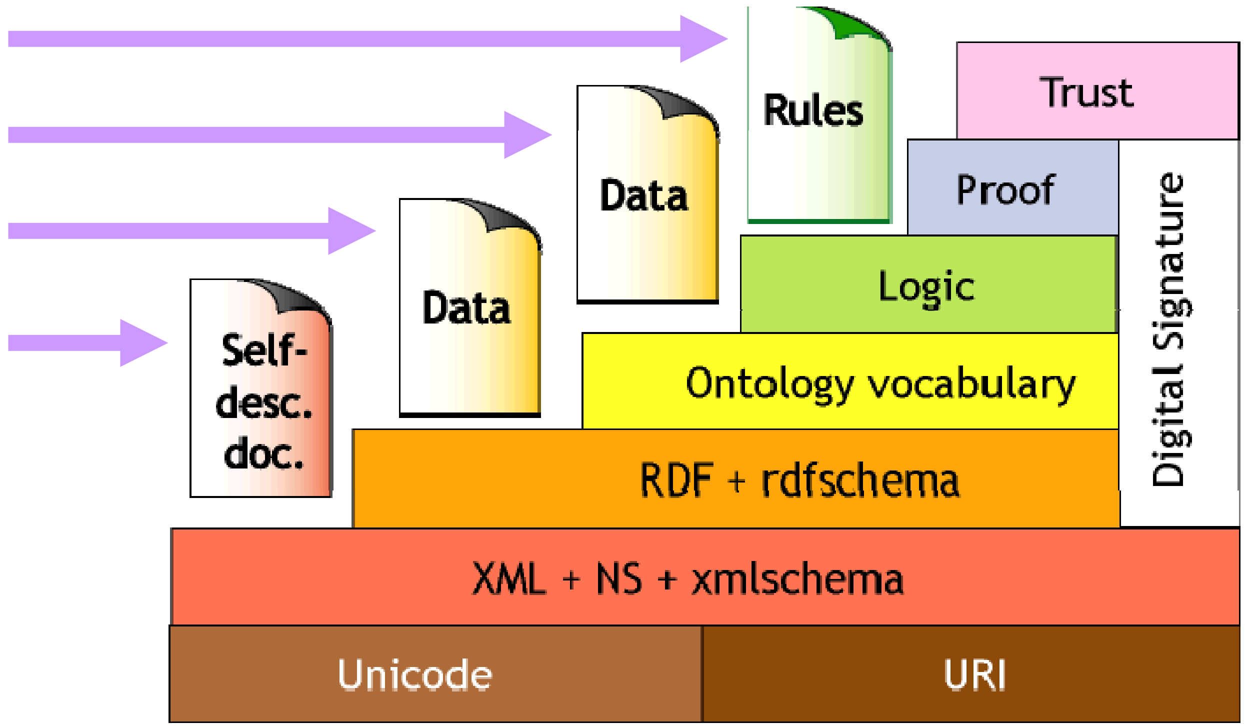
Geospatial + Semantic Web

Geospatial Data
- Geographic info
- Photo imagery
- GIS data
- Maps

- Spatial relations
- Places, landmarks etc.
- Locations, lat/long etc.

- Time
- Temporal relations

- People
- Organizations
- Other things …



(Chen, 2007)

UCSB

Why is this interesting?

- **“Location” is ubiquitous on the Web**
  - Where do you go to school or work?
  - Where did you take your flickr photos?
  - Where is the nearest gas-station from “here”?
  - Where are my friends now?
  - What’s the avg. housing price in my neighborhood?
  - What’s ski condition in MD and PA?
  - …

(Chen, 2007)

UCSB

The present Web is for human

Wenwen LI

Postdoctoral Fellow, Spatial@UCSB
University of California, Santa Barbara
MC4060, 5704 Ellison Hall
Santa Barbara, CA 93106-4060
Phone: 202-615-3835
Email: wenwen at spatial dot ucsb dot edu
UCSB homepage: http://geog.ucsb.edu/~wenwen
GMU homepage: http://mas

Your browser doesn't know that I live in Beijing, China in 2004

Short Bio:

I am a postdoctoral researcher at the Center for Spatial Studies (Spatial@UCSB), Univ Santa Barbara since Aug. 2010. Before joining Spatial@UCSB, I was a research scholar at Center for Intelligent Spatial Computing, College of Science, George Mason University from Feb. 2006 – Feb. 2009. I obtained Ph.D. degree in *Earth System and Geoinformation Science* from College of Science, George Mason University in August, 2010; my M.S. degree in *Signal and Information Processing* from Chinese Academy of Sciences in July 2007 and my B.S. degree in *Computer Science and Technology* from Beijing Normal University in July 2004.

Research Interests:

GIS and Remote Sensing

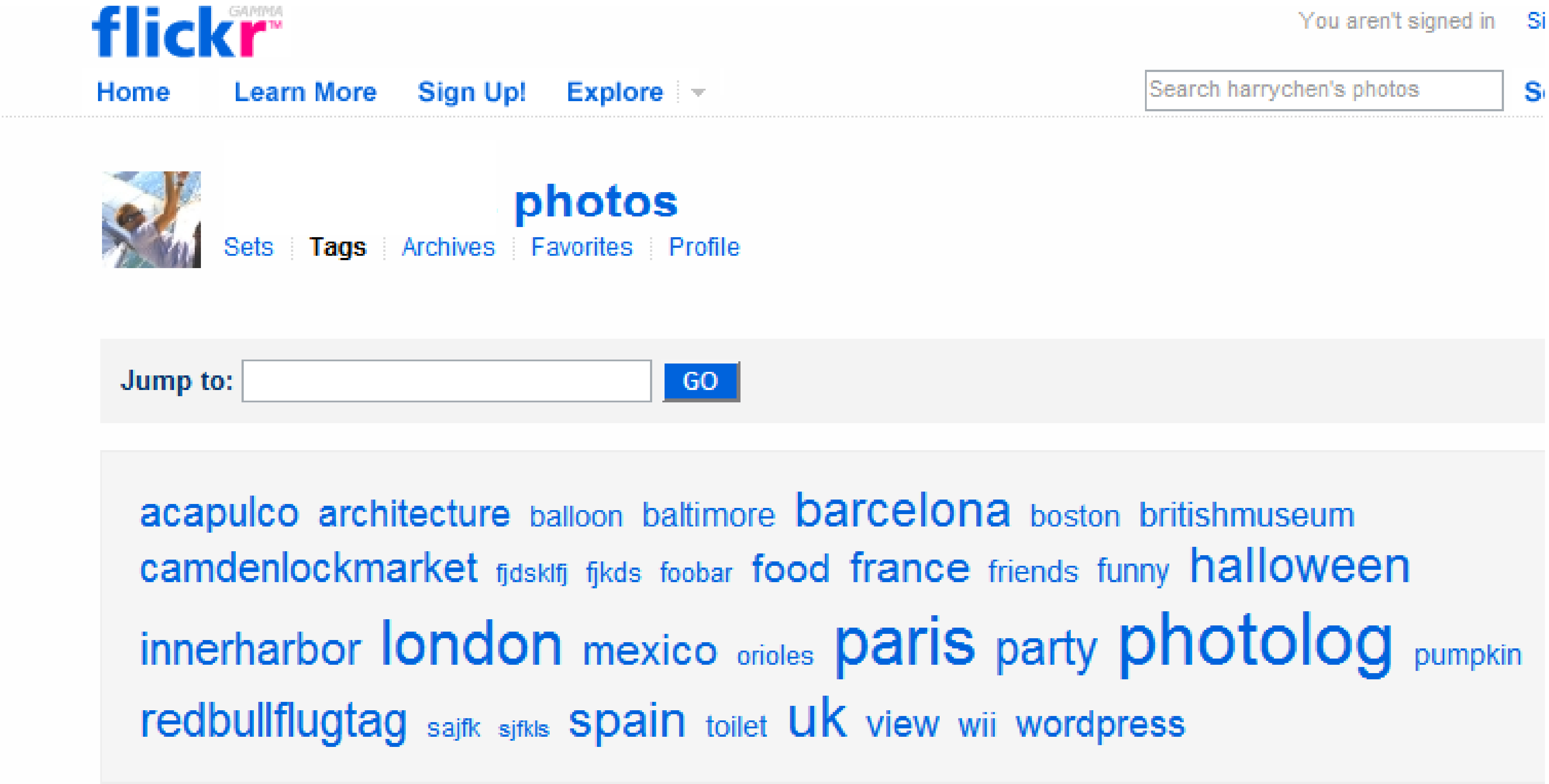
Ontology and Geospatial Semantics: Ontology-aided Semantic Search, Semantic Similarity

Spatial Optimization: Heuristic Modeling

http://geog.ucsb.edu/~wenwen

UCSB

What's "london"?



(Chen, 2007)

UCSB

What's "london" to a machine?

london | United States | search | show on map | [advanced search]

2544 records found for "london"

	Name	Country	Feature class	Latitude	Longitude
1	London Mountain	United States, Colorado	mountain elevation 4018m	N 39° 17' 3''	W 106° 9' 30''
2	New London County	United States, Connecticut	second-order administrative division population 254,957, elevation 22m	N 41° 29' 0''	W 72° 3' 58''
3	New London Nova Londres	United States, Connecticut	populated place population 26,294, elevation 17m	N 41° 21' 20''	W 72° 5' 58''
4	London	United States, Kentucky	populated place population 7,844, elevation 378m	N 37° 7' 44''	W 84° 4' 59''
5	London	United States, Ohio	populated place population 8,771, elevation 321m	N 39° 53' 11''	W 83° 26' 53''
6	New London	United States, Iowa	populated place population 1,857, elevation 232m	N 40° 55' 37''	W 91° 23' 58''
7	New London	United States, Minnesota	populated place population 1,123, elevation 368m	N 45° 18' 3''	W 94° 56' 39''
8	New London	United States, Wisconsin	populated place population 7,043, elevation 234m	N 44° 23' 33''	W 88° 44' 23''
9	London Londres	United States, California	populated place population 1,985, elevation 91m	N 36° 28' 33''	W 119° 26' 35''

http://www.geonames.org/search.html?q=london&country=US

(Chen, 2007)

UCSB

What did we learn?

- Most of the information on the Web today is meant for human consumption.
- Without an explicit semantic description, it’s difficult for machines to consume Web information.
- The study of geospatial semantic web is to exploit Semantic Web and geospatial technology to improve human productivity
  - i.e., get machines to do more work for us.

(Chen, 2007)

UCSB

Motivation

- Vision of digital earth

"..a digital future where schoolchildren - indeed all the world's citizens - could interact with a computer-generated three-dimensional spinning virtual globe and access vast amounts of scientific and cultural information to help them understand the Earth and its human activities."

- Vision of a geoinformatics system

"…a future in which someone can sit at a terminal and have easy access to vast stores of [geoscience] data of almost any kind, with the easy ability to visualize, analyze and model those data."

Challenges

- Huge Amount of Data
  - EOSDIS supports the daily production of over 3 terabytes (TB) of interdisciplinary Earth system science data.
- Widely Dispersed
  - DAAC, SEDAC,MODAPS,DISC, ASDC…
- Poorly Catalogued
  - GOS: 15K layers, 600 WMS, only 80 are live
- Difficult to Interoperate
  - Different vendors, various formats

Needs

- An effective data sharing infrastructure that data providers can advertise their data to make them publicly visible.
- A common sense knowledge framework that machine agents can understand not only the syntax but also the semantics of the data.
- A mechanism for seamless integration of geospatial resources.

Goals

- **Sharing**

  Earth observation data and services

- **Sharing**

  - Sharing software architecture, reusable components and technical solutions.

- **Sharing**

  Public:

  - fast access and acquisition of Earth Science Data

  Professionals:

  - mechanism for data integration
  - efficient data management

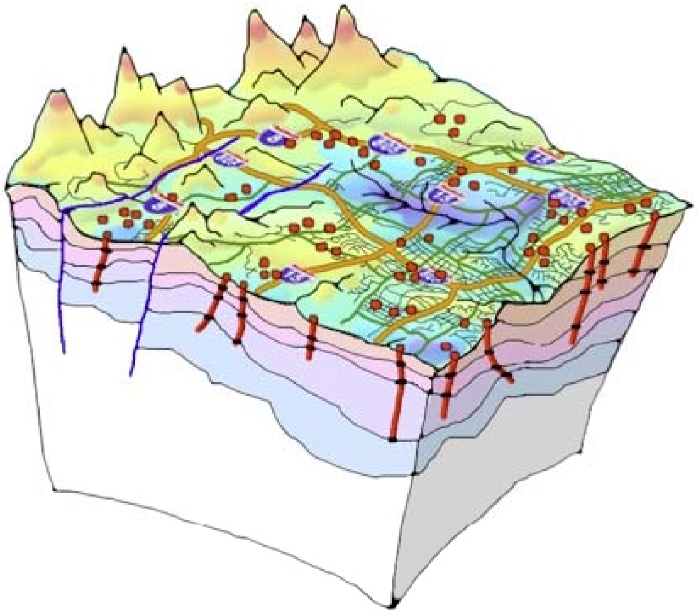
  Decision makers:

  - models and accurate predictions for future decision making.

UCSB

The GEON Example

"For a given region (i.e. lat/long extent, plus depth), return a 3D structural model with accompanying physical parameters of density, seismic velocities, geochemistry, and geologic ages, using a cell size of 10km"



OpenEarth Framework (OEF)

Data Types

- Standard DEM data, satellite imagery, street maps, geologic maps and other coverage data.
- Geophysical data: seismic, gravity and magnetic data.
- Bore hole or well data: rock types

UCSB

The VASDI Example

"How does melting snow and sea ice influence habitat changes of polar wildlife?"

<u>Spatial and Temporal data</u>

- ✓ **Habitat patch for polar bear**
- ✓ **Snow concentration**
- ✓ **Sea ice concentration**

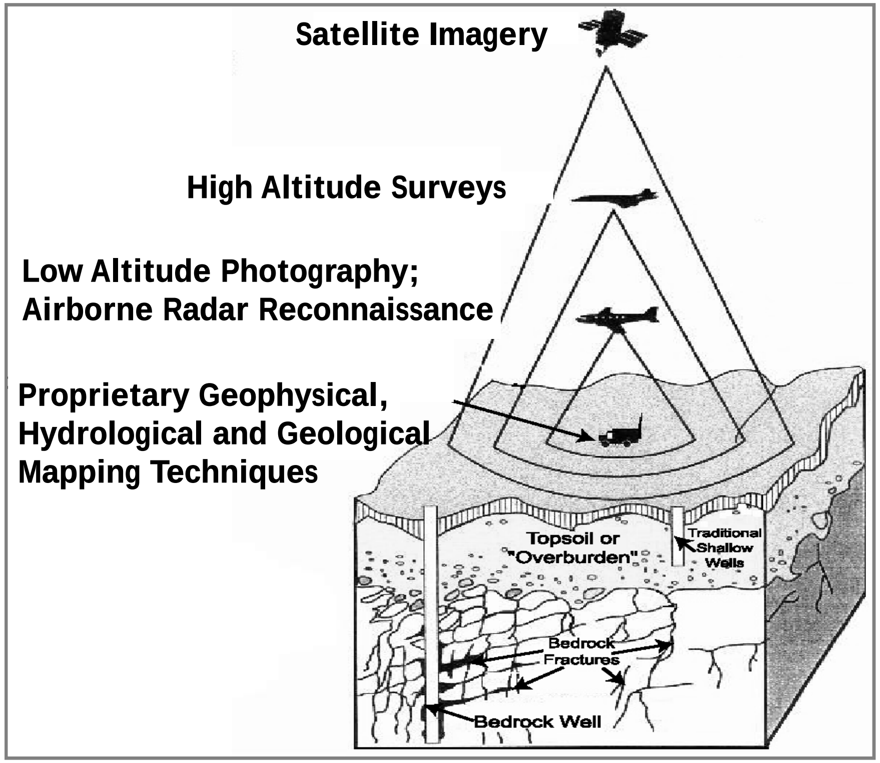
Model

- ✓ **Statistical model**

UCSB

The Example of Hydrological Modeling

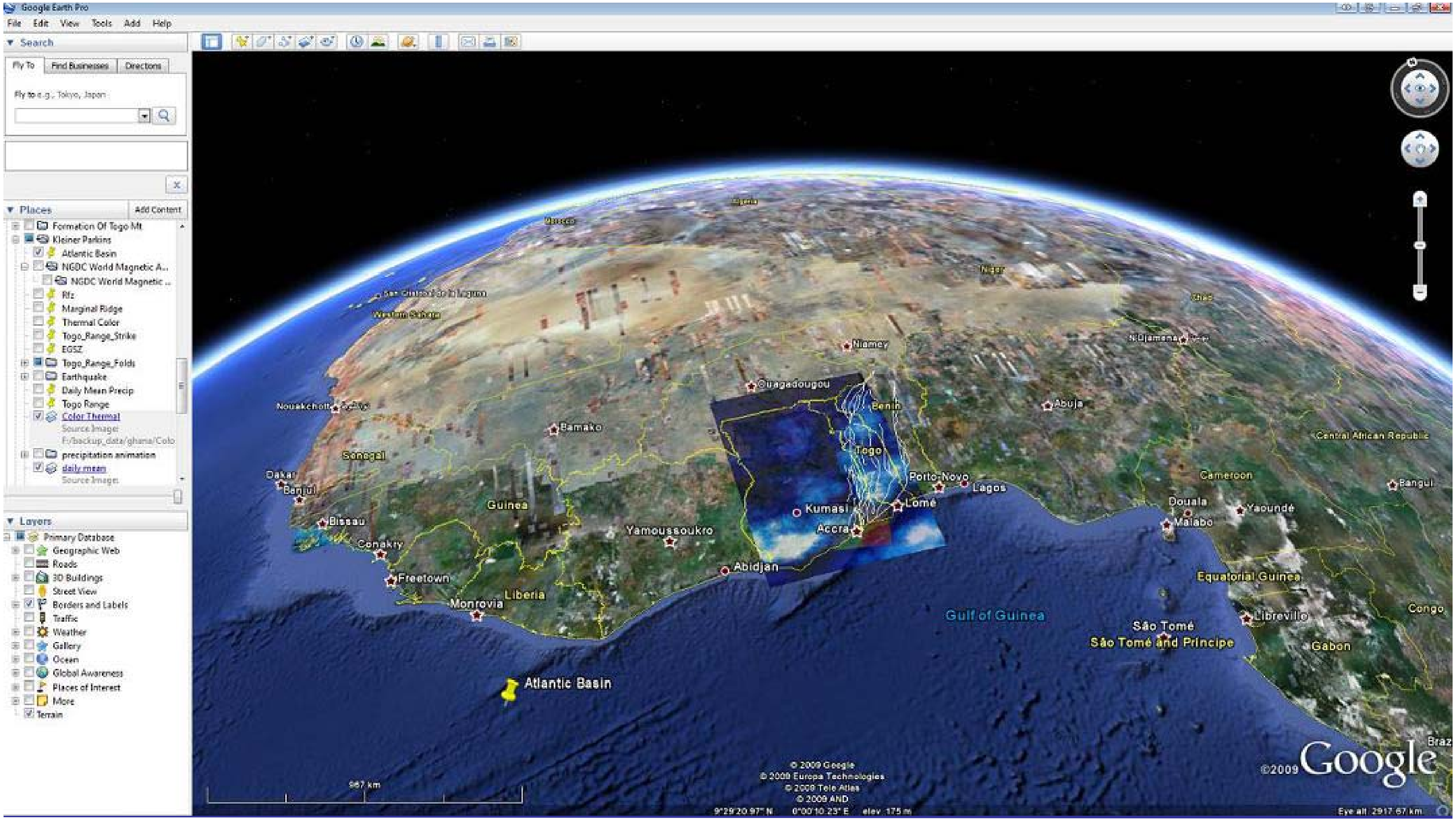
Exploring sustainable groundwater resources to resolve the Global Water Crisis– The case of Ghana



megawatershed

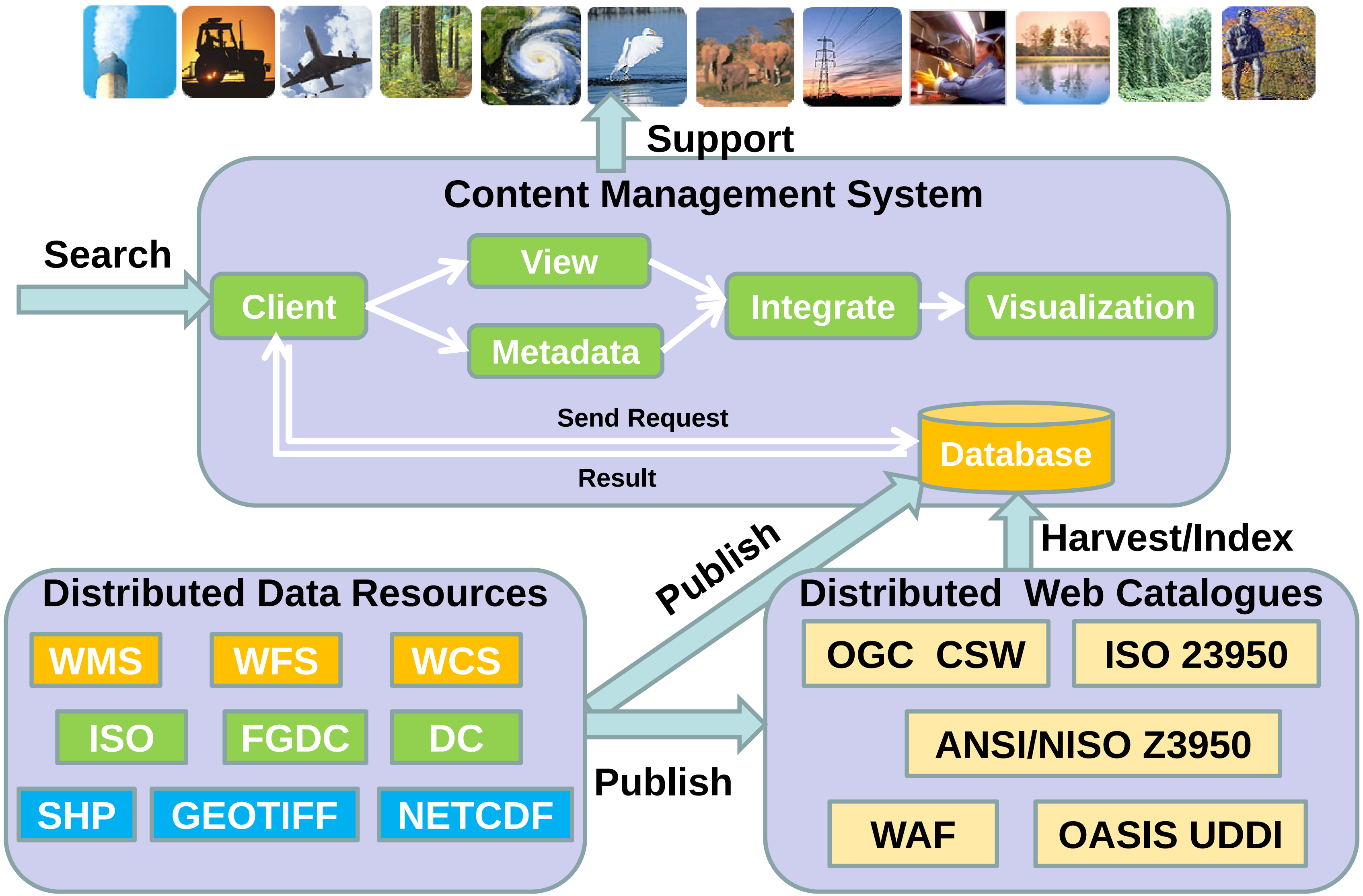
UCSB

The Example of Hydrological Modeling



UCSB

Geospatial CyberInfrastructure



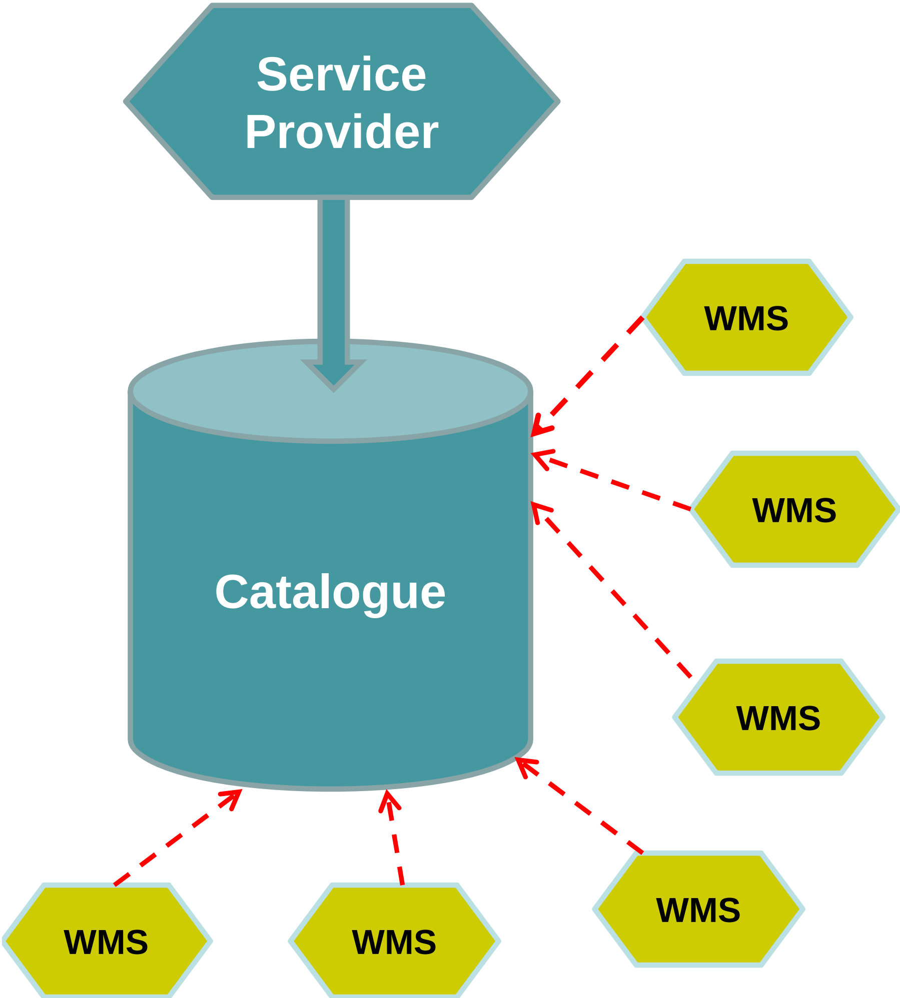
Sharing of Geoscientific Data

- OGC (Open Geospatial Consoritum)
- Data Standards
  - Web Map Service (WMS)
  - Web Feature Service (WFS)
- Catalog Standard
  - Catalog Service for the Web (CSW)

Geospatial Web Service (GWS)

A Hybrid Approach for OGC GWS Discovery

Existing Methods

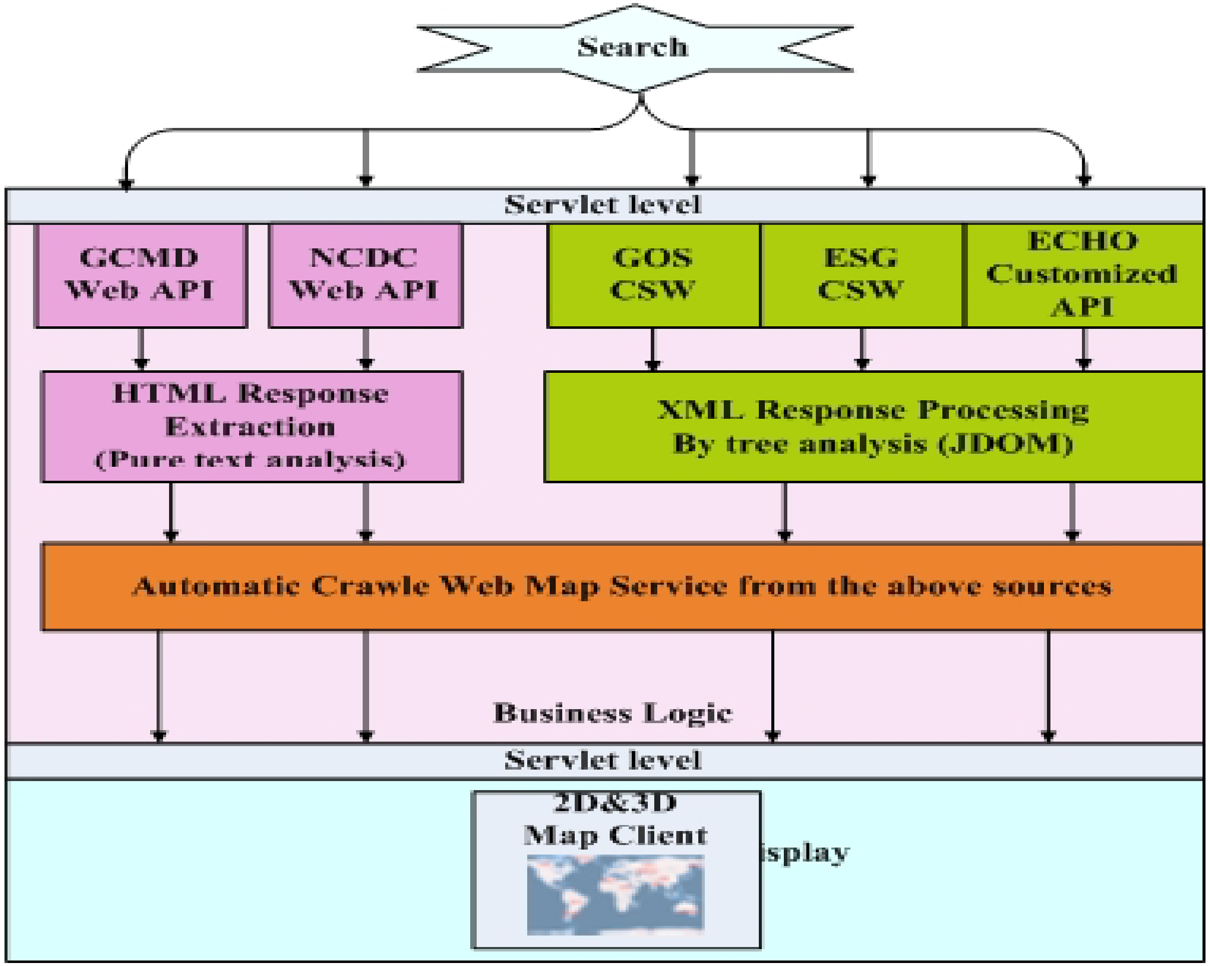


- OGC CSW Catalogs
  - NASA GCMD
  - USGS GOS
  - NOAA NCDC
  - NASA ECHO
  - UC INSPIRE
- Disadvantages
  - *Lack of update* (Ma 2004)
    - Don't update in a timely fashion.
  - *Passive mode* (Al-Masri 2007)
    - Requires manual registration.

UCSB

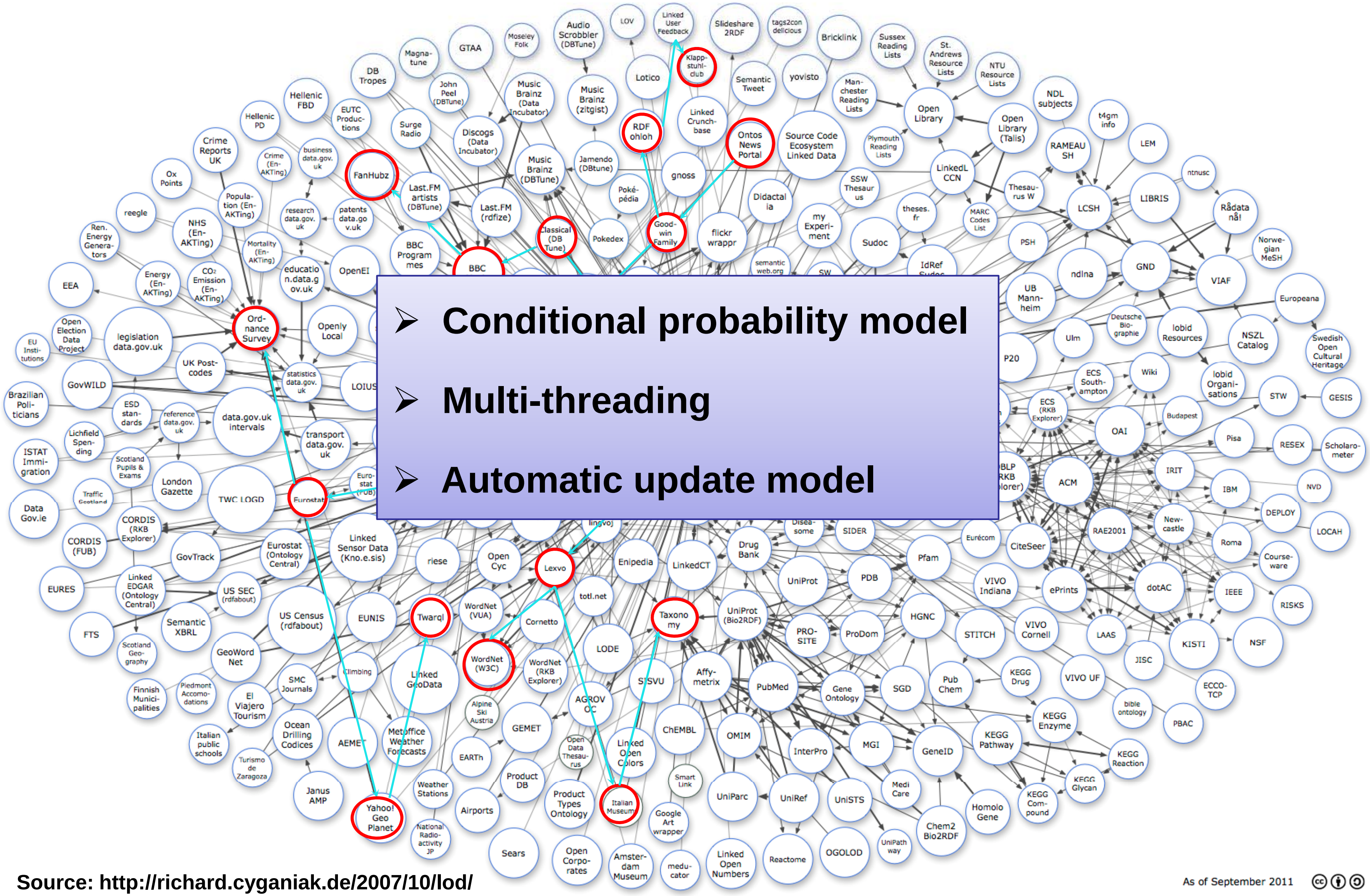
Geo-bridge for Meta-Catalogue

- Standards:
  - Web Catalogue Service (CSW) – GOS, ESG
  - Customized API – ECHO
  - Web Interface – GCMD, NCDC
- Seamless Communication
  - XML-encapsulated Request
  - KVP-based Request
- Service Parser
  - HTML parser
  - XML parser
- Key Techniques
  - Ajax: Asynchronous JavaScript and XML
  - Multi-Threads



UCSB

Distributed Crawler

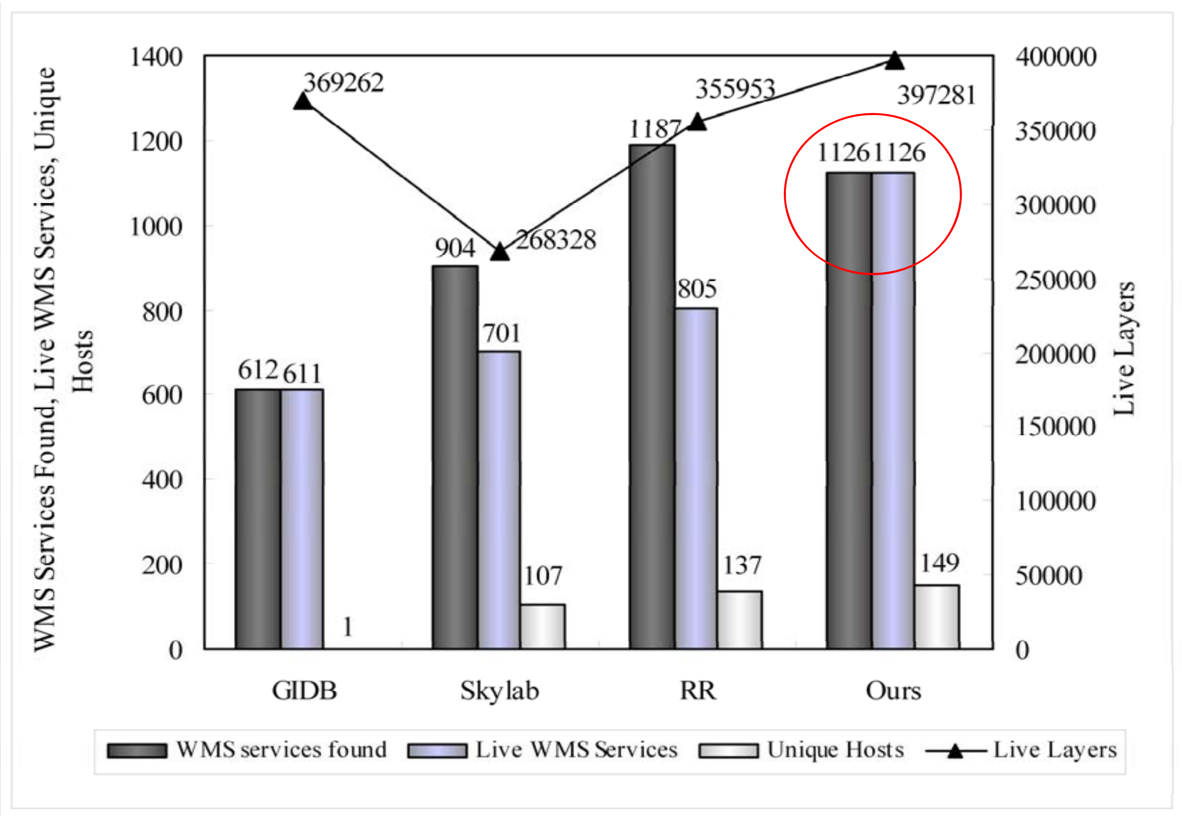


- Conditional probability model
- Multi-threading
- Automatic update model

Source: http://richard.cyganiak.de/2007/10/lod/

Experimental Results – Scalability Test

- Scalability test



UCSB

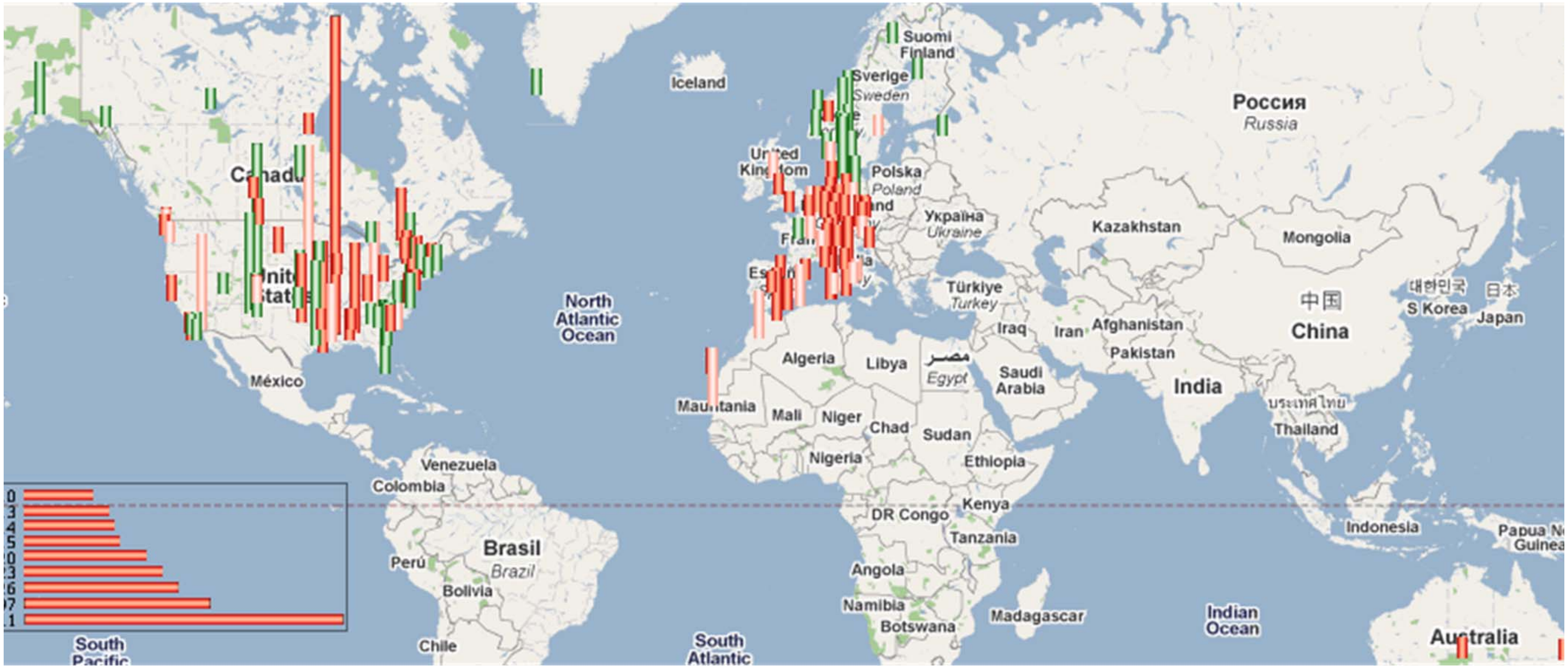
Dynamics in Space & Time

- Dec. 2008 (1126 WMS; 18 countries)

UCSB

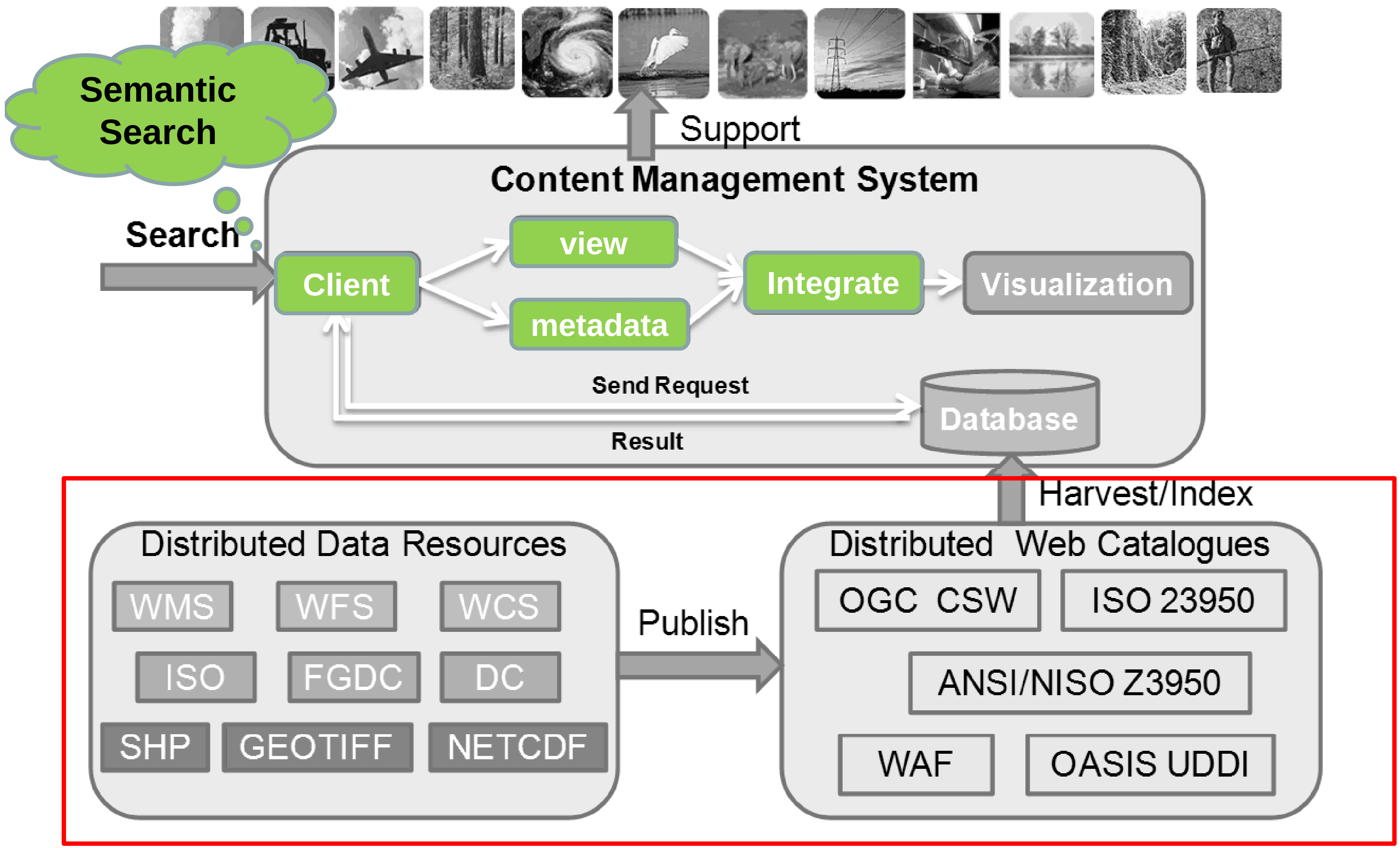
Dynamics in Space & Time

- June 2009 (+228 WMS; -168 WMS)



UCSB

Geospatial CyberInfrastructure



Ontology-based Knowledge Discovery and Integration

Geo-Ontologies

- Definition
  - Classes and individuals for representing e.g. geospatial objects, their properties, and mutual relationships.

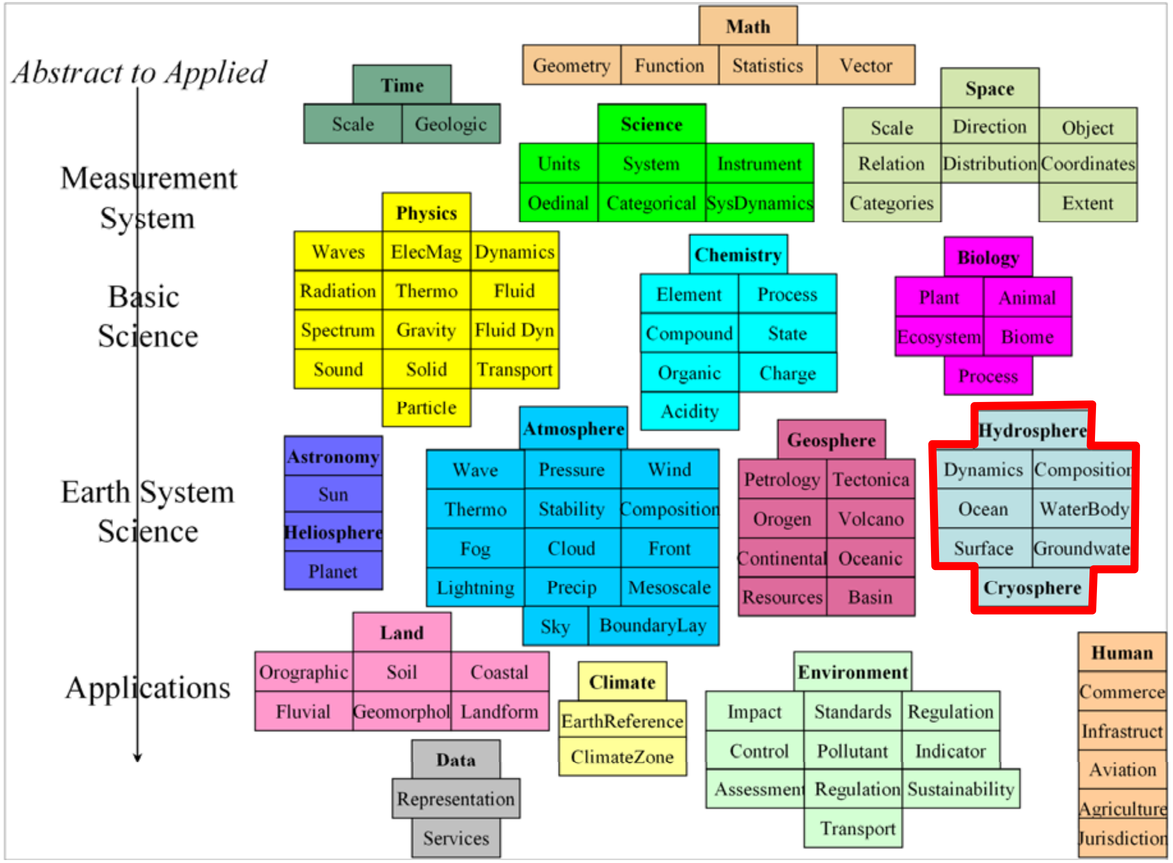
- Ontology
  - Plays an important role in establishing robust theoretical foundations for GIScience in the future (David Mark)

SWEET 2.0

INSPIRE Ontology

NASA GCMD Science Keyword

Dbpedia Geolinked data



GEMET Thesaurus

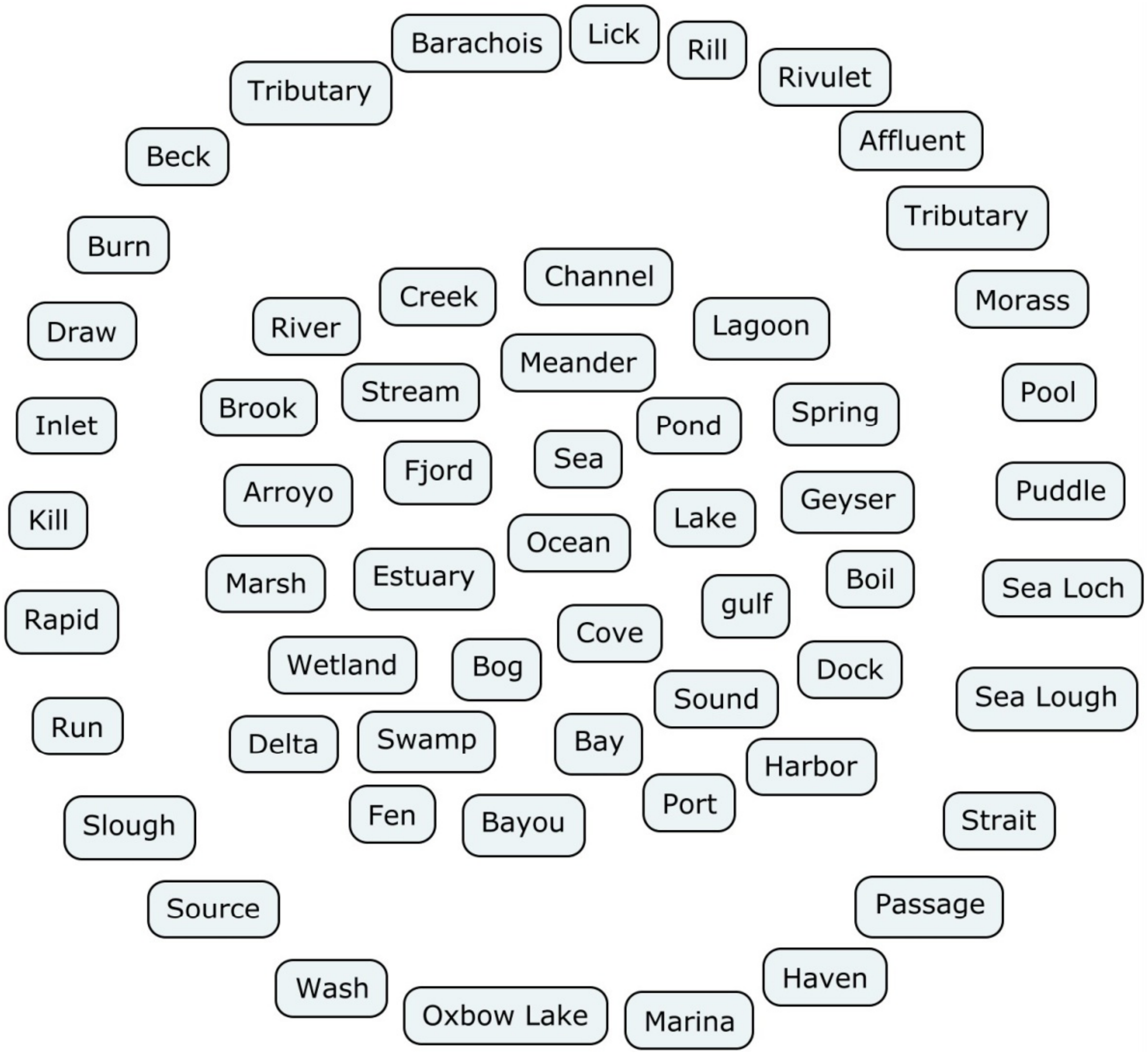
CUASHI Ontology

MMI Ontology

SWEET: Semantic Web for Earth and Environmental Terminology

UCSB

Waterbody Ontology >> Object Space



Waterbody Ontology >> Attribute Space

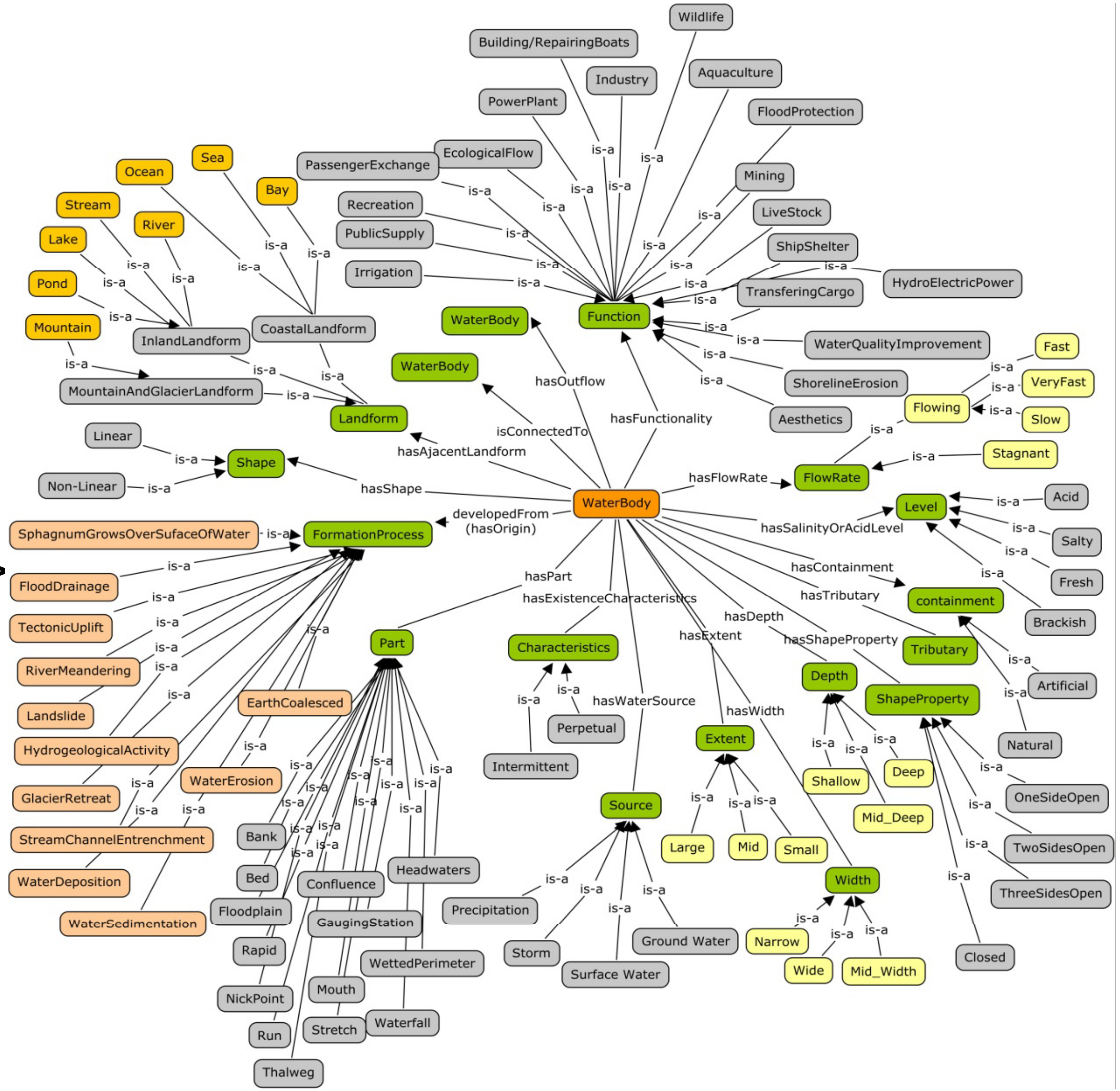
- ✓ hasFormationProcess
- ✓ hasPart
- ✓ hasExistenceCharacteristi
- ✓ hasWaterSource
- ✓ hasExtent
- ✓ hasWidth
- ✓ hasDepth
- ✓ hasShape
- ✓ hasShapeProperties
- ✓ hasTributary
- ✓ hasContainment
- ✓ hasSalinityLevel
- ✓ hasFlowRate
- ✓ hasFunction
- ✓ hasOutflow
- ✓ isConntectedTo
- ✓ hasAjacentLandform

UCSB

Waterbody Ontology >> Value Space

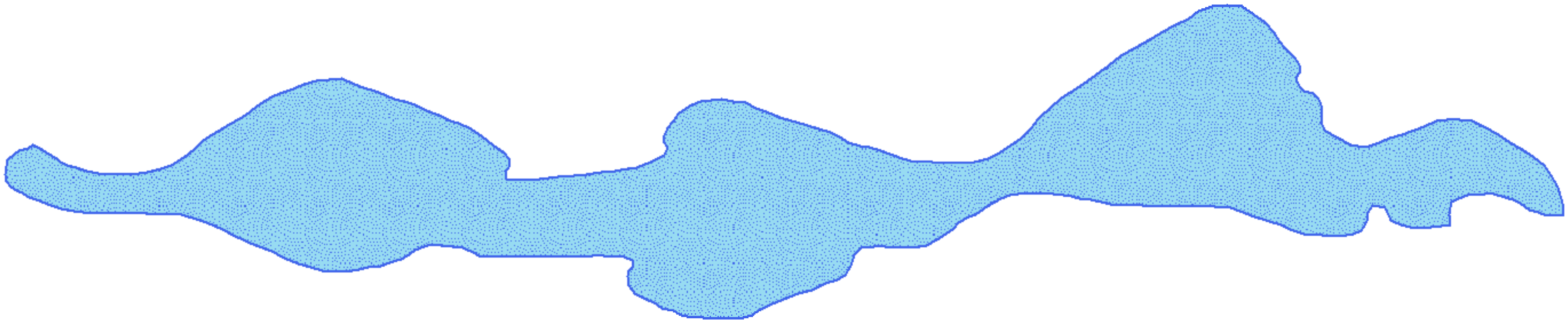
- ✓ <River, hasShape, linear>
- ✓ <River, hasFlowRate, flowing>

- ✓ <Lake, hasShape, nonlinear>
- ✓ <Lake, hasFlowRate, stagnant>



Similarity Reasoning

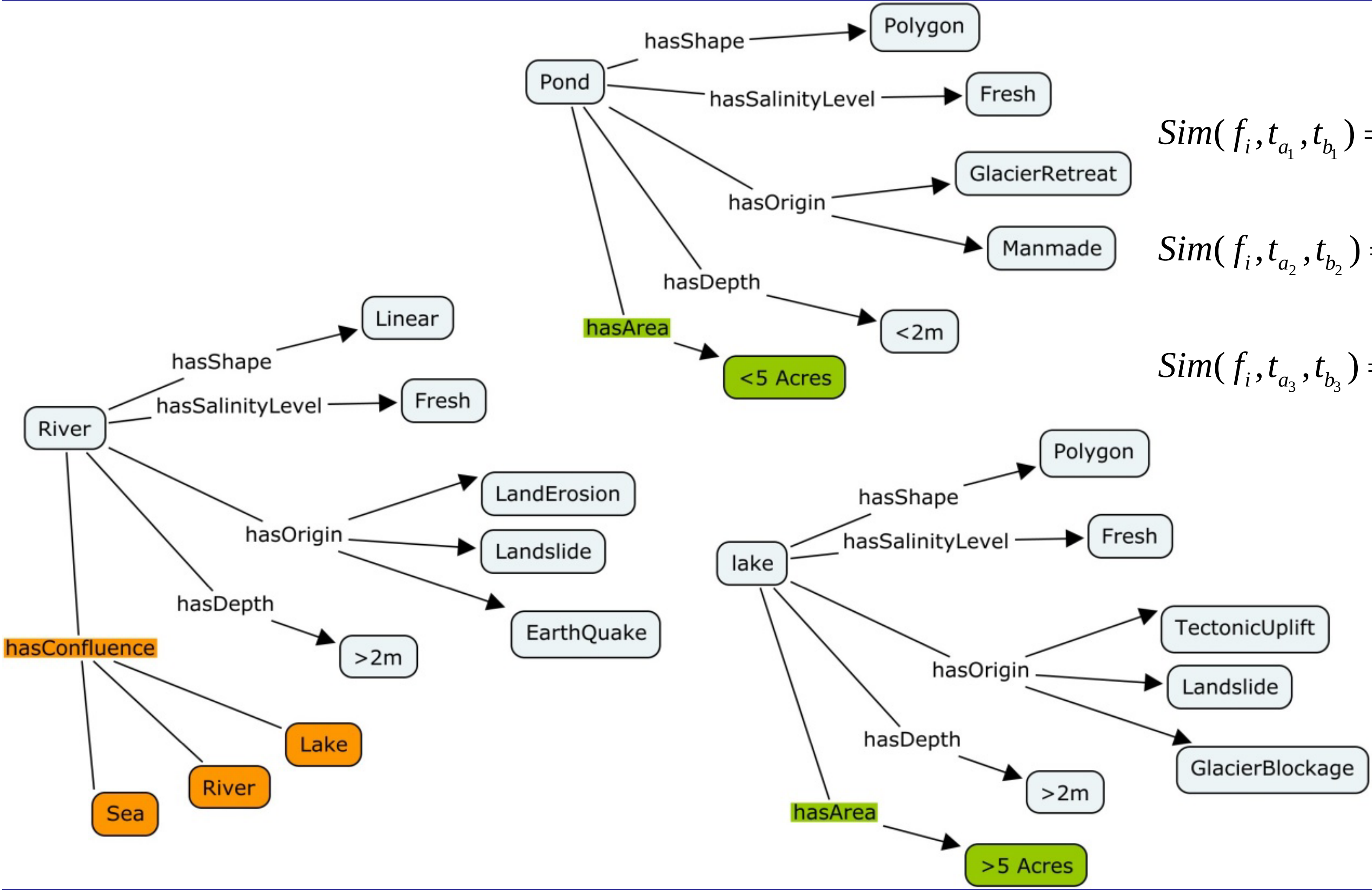
- Objective
  - Eliminate the intrinsic vagueness
  - Identify objects conceptually close



Vagueness in water features: three lakes or a meandering river? (Santos et al. 2005)

UCSB

Similarity Reasoning



$$Sim(f_i, t_{a_1}, t_{b_1}) = \alpha \frac{1}{|A_1 \cup B_1|};$$

$$Sim(f_i, t_{a_2}, t_{b_2}) = \alpha Sim(f_i, t_{a_1}, t_{b_1});$$

$$Sim(f_i, t_{a_3}, t_{b_3}) = \frac{|A_3 \cap B_3|}{|A_3 \cup B_3|}$$

UCSB

Logic Reasoning (1)

Case study: "***How does solid water melt influence stream flow in the Arctic Region over the summer time?***"

- Syntax Analysis – Query decomposition
  - Component
    - What- "Solid water"
    - Place - "Arctic"
    - Process – "Change"
    - Time – "Summer"
  - Description logic-based query

Q1: $\text{Solid Water} \sqcap \exists \text{hasProperty.Change} \sqcap \exists \text{hasObject.Stream} \sqcap \forall \text{takePlaceIn.Arctic} \sqcap \forall \text{hasTime.Summer}$

Logic Reasoning (2)

- Semantic Analysis

Q1a: $\textbf{SomeSWClass} \cap \exists isSubClassesOf.\text{"Solid Water"}$

Q1b: $(\text{AProperty} \cap \exists isSubClassOf.\text{"Property"})$

$\cap (\text{AProperty} \cap \exists is\,\text{Pr}\,edicateOf.\text{SomeSWClass})$

$\cap (\textbf{Parameter} \cap \exists isObjectOf.\text{SomeSWClass})$

Q1c: $\textbf{SomeStreamClass} \cap \exists isSubClassesOf.\text{"Stream"}$

Q1d: $(\text{Parameter} \cup \text{SomeStreamClass}).\,\text{hasData} \cap \forall takePlaceIn.\text{"Arctic"} \forall hasTime.\text{"Summer"}$

UCSB

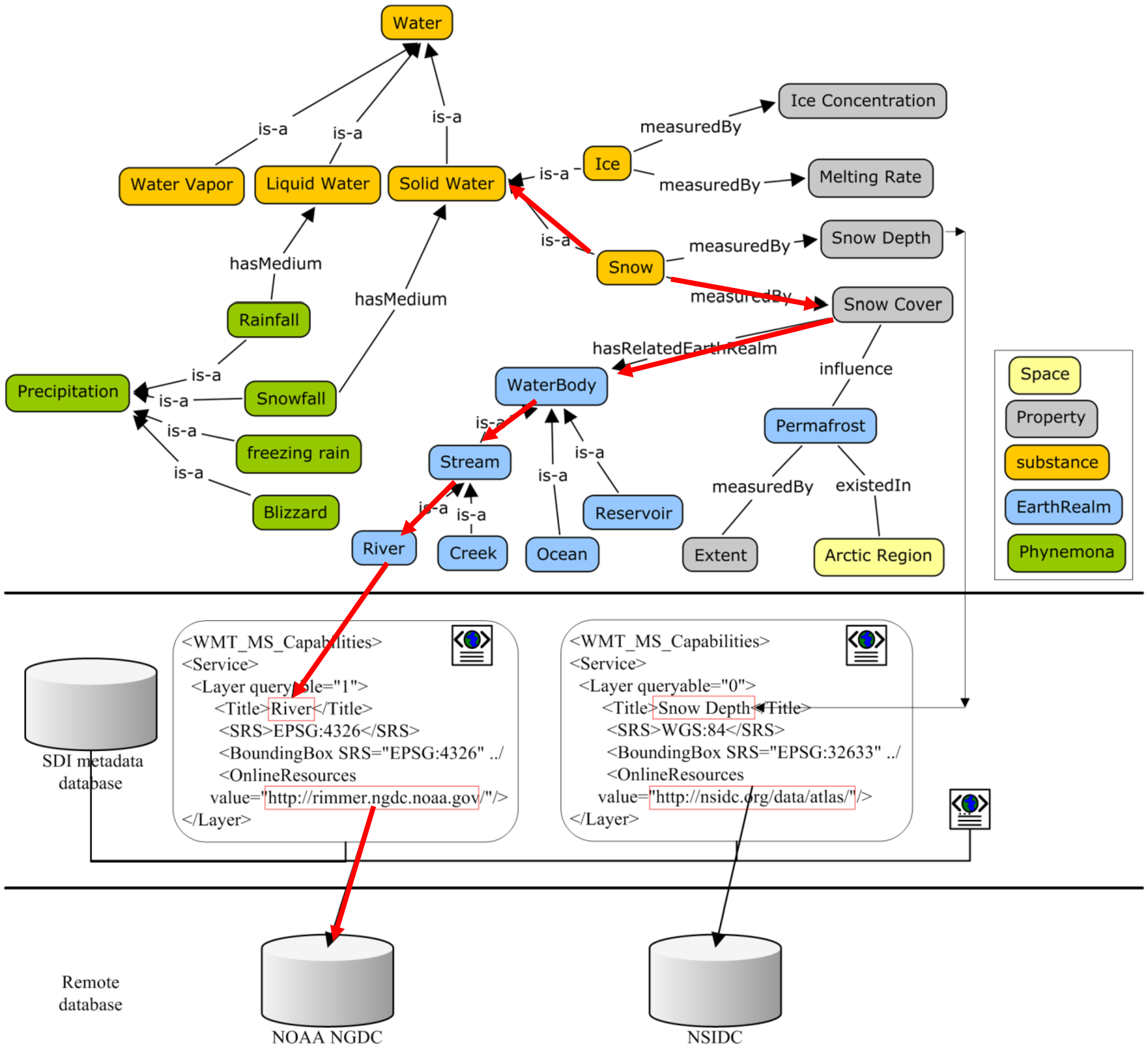
Logic Reasoning (3)

- Formal Query – Machine language
  - Q1d

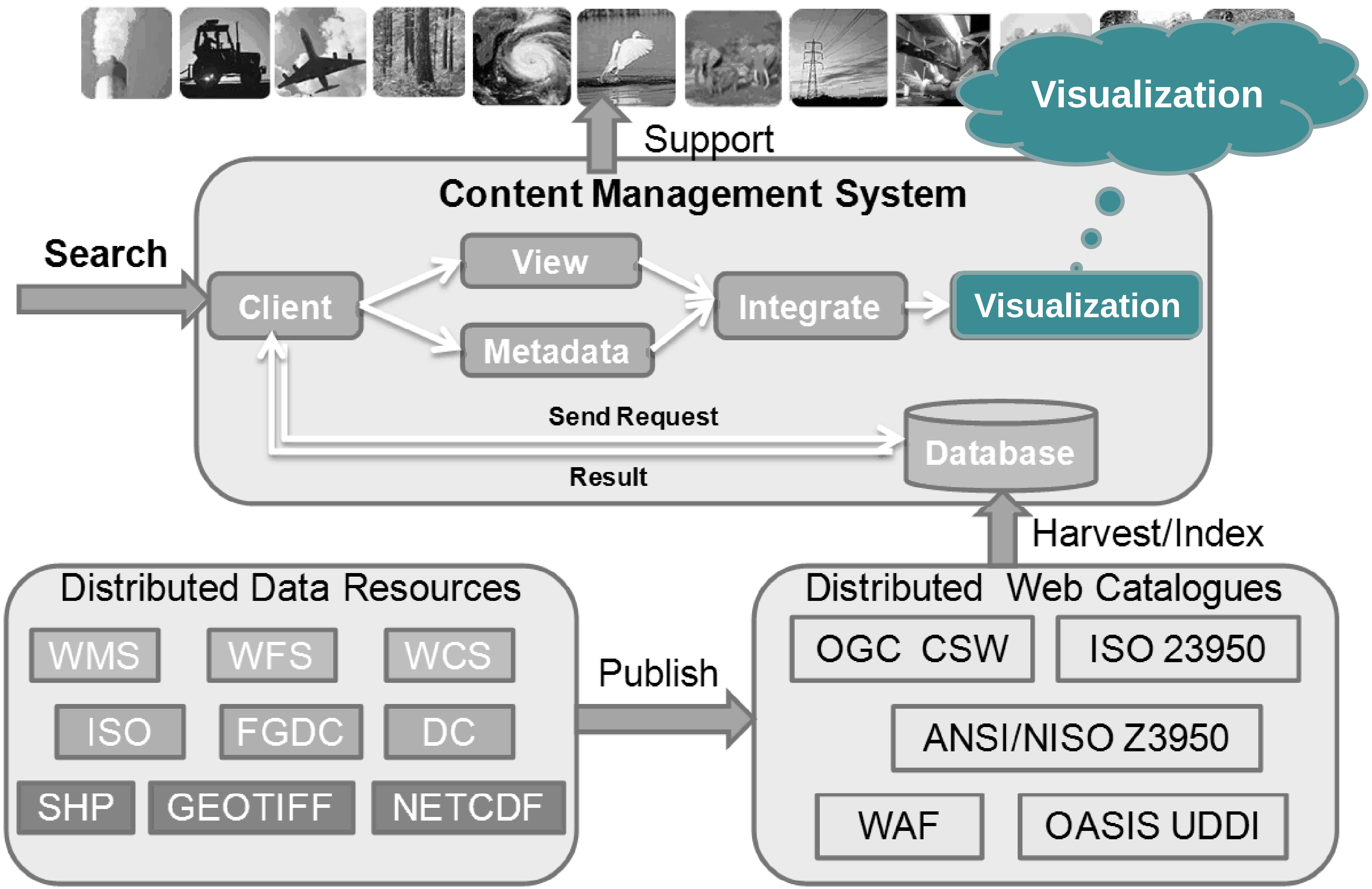
```
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:  <http://www.w3.org/2002/07/owl#>
PREFIX PhenomenaNS: <http://localhost/ontology/phenomena.owl#>
PREFIX PropertyNS: <http://localhost/ontology/property.owl#>
PREFIX SubstanceNS: <http://localhost/ontology/substance.owl#>
PREFIX EarthRealmNS: <http://localhost/ontology/earthrealm.owl#>
PREFIX ProcessNS: <http://localhost/ontology/process.owl#>

SELECT *
     WHERE {
    ?Parameter PropertyNS:hasData ?data
    ?data PropertyNS:takePlaceIn ‘Arctic’^^xsd:String
    ?data PropertyNS:hasTime ‘Summer’^^xsd:String
}
```

UCSB

Logic Reasoning to Answer Science Question



Geospatial CyberInfrastructure

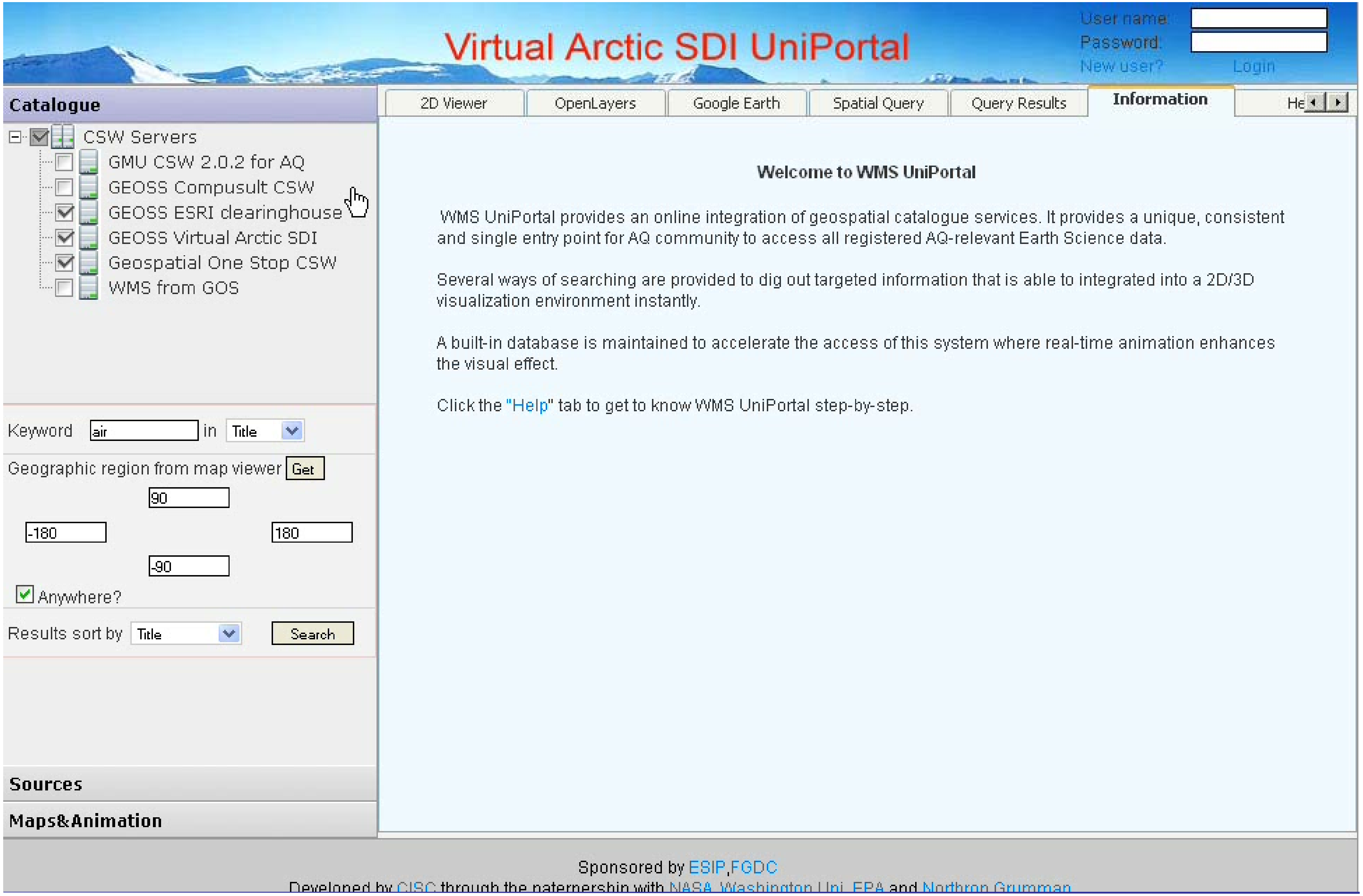


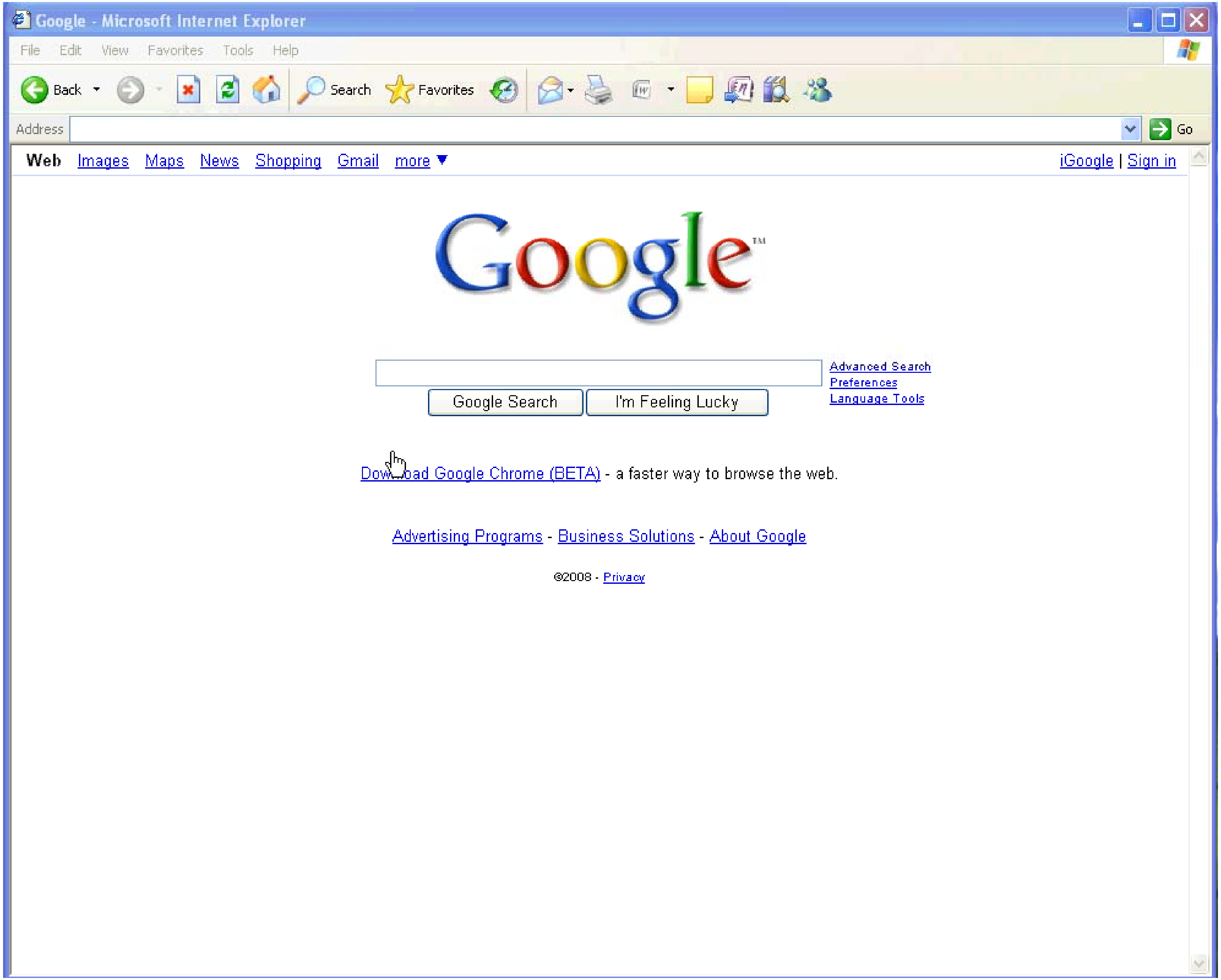
Open Source Visualization Tools

- 2D
  - Openlayers (http://openlayers.org/)
  - OGC Viewer (http://www.wmsviewer.com/)
  - QuickWMS (http://inovagis.terradue.com/quickwms/index.htm)
- 3D
  - Microsoft Virtual Earth
  - Google Earth
  - ESRI ArcGlobe
- 4D
  - 3D+Time

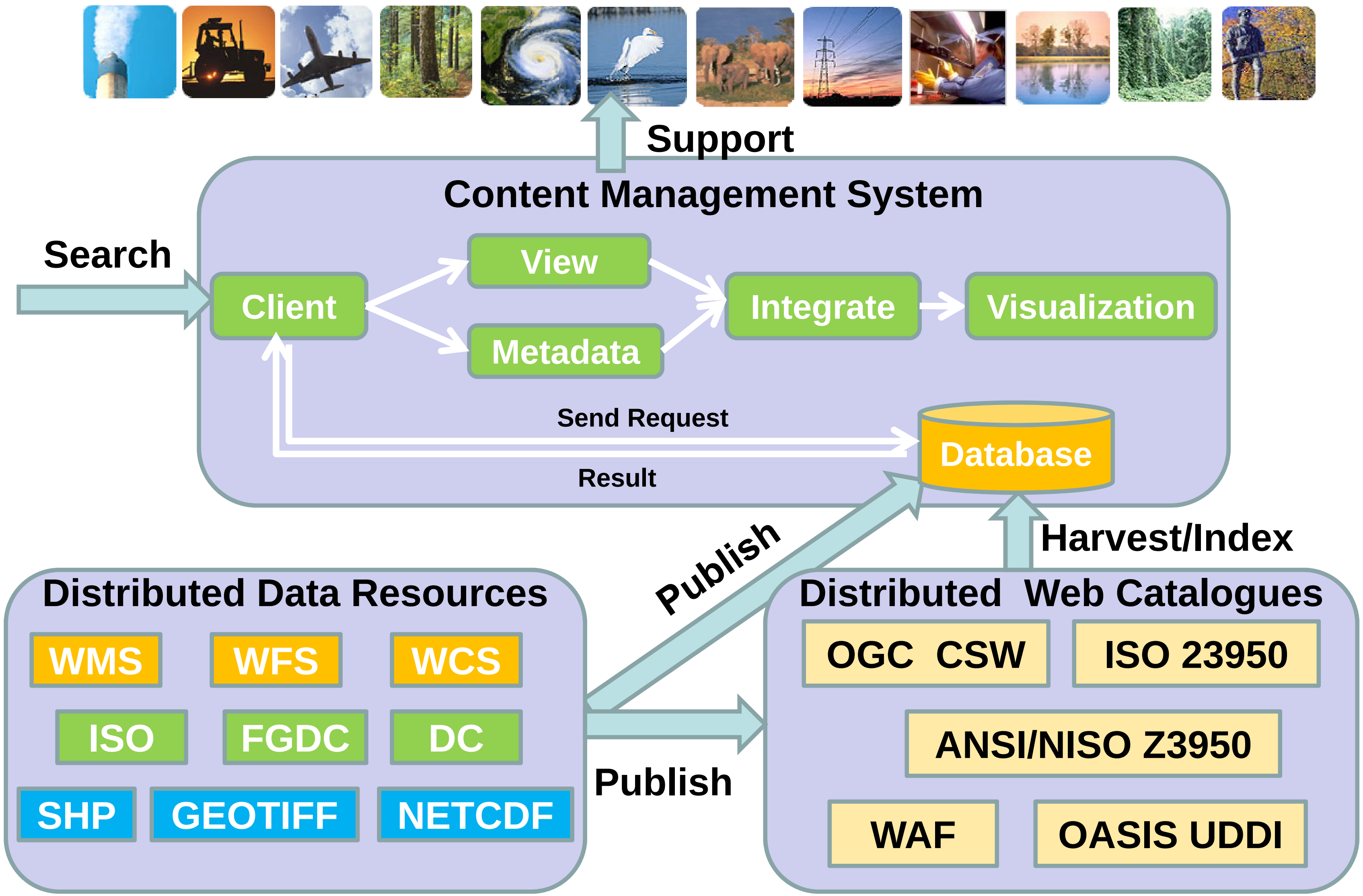
Applications and Demos

- Arctic Spatial Data Infrastructure
- Hydrological Semantic Search


Virtual Arctic SDI UniPortal
User name:
Password:
New user?
Login
Catalogue
2D Viewer
OpenLayers
Google Earth
Spatial Query
Query Results
Information
He
CSW Servers
GMU CSW 2.0.2 for AQ
GEOSS Compusult CSW
GEOSS ESRI clearinghouse
GEOSS Virtual Arctic SDI
Geospatial One Stop CSW
WMS from GOS
Welcome to WMS UniPortal
WMS UniPortal provides an online integration of geospatial catalogue services. It provides a unique, consistent and single entry point for AQ community to access all registered AQ-relevant Earth Science data.
Several ways of searching are provided to dig out targeted information that is able to integrated into a 2D/3D visualization environment instantly.
A built-in database is maintained to accelerate the access of this system where real-time animation enhances the visual effect.
Click the "Help" tab to get to know WMS UniPortal step-by-step.
Keyword
air
in
Title
Geographic region from map viewer
Get
90
-180
180
-90
Anywhere?
Results sort by
Title
Search
Sources
Maps&Animation
Sponsored by ESIP,FGDC
Developed by CISC through the paternership with NASA, Washington Uni, EPA and Northrop Grumman


Google - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Address
Go
Web Images Maps News Shopping Gmail more
iGoogle | Sign in
Google
Advanced Search
Preferences
Language Tools
Google Search
I'm Feeling Lucky
Download Google Chrome (BETA) - a faster way to browse the web.
Advertising Programs - Business Solutions - About Google
©2008 - Privacy

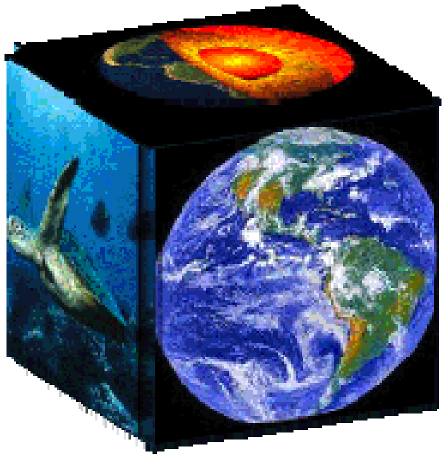
Summary


Support
Content Management System
Search
Client
View
Metadata
Integrate
Visualization
Send Request
Result
Database
Publish
Harvest/Index
Distributed Data Resources
WMS
WFS
WCS
ISO
FGDC
DC
SHP
GEOTIFF
NETCDF
Publish
Distributed Web Catalogues
OGC CSW
ISO 23950
ANSI/NISO Z3950
WAF
OASIS UDDI

Future Work

Observation	Understanding	Models	Prediction	Consequences
Data Proficiency	Assessment Synthesis Analysis	Process Coupling Validation	Assimilation Diagnosis Prognosis	Applications Education Linkages

Cyberinfrastructure ⟷ **Science**



NSF EarthCube

- Thanks & Questions